This tutorial will provide a general outline on how to optimize a molecular model. Validation will follow in the next tutorial.

The goal of this tutorial will be to optimize a published crystal structure (PDBid: 1DP0) with the 3.2Å beta-galactosidase density map (EMD-5995). This model was optimized manually using COOT in 2014, and the resulting model was deposited in the PDB (3J7H). Ideally, our real-space refinement method of the model will produce a better model (both stereochemistry and fit-to-density). Moreover, we look to accomplish this within the 1 hour tutorial. The validation tutorial that follows will allow us to determine if an improvement was made. Finally, before going to break we will optimize our reconstructed EMAN2 result which will most likely be slightly worse in resolution compared to the 3.2Å data.

---

Before we start, here is some helpful information:

```
phenix.real_space_refine: tool for extensive real-space refinement
of atomic coordinates against provided map

Usage:
  phenix.real_space_refine model.pdb ccp4_formatted_map.map
  or
  phenix.real_space_refine model.pdb map.mtz
  or
  phenix.real_space_refine model.pdb map.mtz
label='2FOFCWT,PH2FOFCWT'
```

or use the following to obtain all parameters.

```
phenix.real_space_refine --h
```

---

## Download / Preprocess the Data

1. Through Chimera (**File > Fetch by ID…**), download the PDB: 1DP0, and the EMDB map: 5995. OR open the following files with Chimera (**File>Open…**), in the model optimization folder (files include 1DP0.pdb and emd_5995.map).
2. Map fitting process before optimization
   ***** The following steps (Step 2 and 3) are not needed during the tutorial since the data is provided. This is a good outline of exactly what I did to prepare the data. *****
   2.1. As one can see the file are not oriented the same due to being derived from different data sets. We will need to first align the data sets before optimizing. To do this open the model panel (**Favorites>Model Panel**), select the model from the list on the left and click the **activate only button** on the right. This

allows the user to manually move the model into the density map. Once moved into the map press activate all.

2.2. Open the command line (**Favorites>Command Line**) and type "**fit #0 #1**", where #0 is the model and #1 is the map. The model should now be fit, however this is not the optimal model for the density map.

2.3. The user can then adjust the map in volume viewer to improve visualizing the data. To do this open the volume viewer (**Tools> Volume Data> Volume Viewer**) and change the step box to 1. In addition the user can adjust the threshold.

2.4. At this point we want to select one chain to optimize. This will allow us to quickly optimize a portion of the complex with a large amount of flexibility, and quickly. To do this click (**Select>Chain A**) followed by (**Select>Invert (Selected Models)**). This will allow us to select everything but Chain A. It should be noted that this can be done though the command line.

2.5. Remove the selected atoms by pressing (**Actions>Atoms/Bonds>Delete**).

2.6. Next, we want to remove non-protein atoms in the model (Waters for instance). To do this press (**Select>Residue>all nonstandard**) and remove with (**Actions>Atoms/Bonds>Delete**).

2.7. The resulting model needs to be saved. To do this, press (**File>Save PDB…**). Save the pdb as **1DP0_chainA.pdb** and be sure to **save relative to the map** (checkbox below). *VERY IMPORTANT*

3. We now have a fit subunit to operate on. So we open a terminal and navigate to the directory with the map file and the PDB file.

Once there we will run the following:

```
Optimize hryc$ phenix.real_space_refine 1DP0_chainA.pdb
emd_5995.map resolution=3.2
```

3.1. and press enter.

3.2. An **error** should occur. But this is an informative error:

```
1DP0_chainA.pdb (149.6, 168.38, 200.69, 90, 90, 90) P 21
21 21
emd_5995.map (216.75, 216.75, 216.75, 90, 90, 90) P 1
Sorry: Crystal symmetry mismatch between different
files.
```

This indicates that we will need to change the **CRYST** entry in the pdb to match the new density map.

3.3. To do this we will open the pdb file 1DP0_chainA.pdb in a text editor.

1.3.1. Change the CRYST1 file to match the unit cell size of the map and P1 symmetry.

1.3.2. In addition, remove all non-atom, CRYST1, and END lines in the PDB file.

3.4. Feel free to open the 1DP0_chainA.pdb file in a text editor to see what the resulting PDB should look like.

4. Good to go!

## Run Initial Phenix.real_space_refine

At this point we now have a single subunit ready for a quick optimization. At this step we like to see the most amount of movement in the optimization process. We can adjust parameters to provide this movement, however, more movement requires more time. Before running Phenix.real_space_refine we can see what is needed by typing in the following to the terminal:

```
Optimize hryc$ phenix.real_space_refine
```

To see additional parameters, type the following to the command line:

```
Optimize hryc$ phenix.real_space_refine --h
```

1. Now let's run Phenix.real_space_refine on the single subunit (with altered model) using default parameters.

   ```
   Optimize hryc$ phenix.real_space_refine 1DP0_chainA.pdb
   emd_5995.map resolution=3.2
   ```

   Various start information should appear including some basic model statistics. Ideally, the model will converge to an improved model. With one subunit, and default parameters, the model should converge to the result quickly. This should take between 3 and 6 minutes.
2. The result should include the following (and will be written into the same directory):
   a. 1DP0_chainA_initial.geo - geometry restraint file
   b. 1DP0_chainA_real_space_refined_all_states.pdb - all intermediate models including the converged model.
   c. 1DP0_chainA_real_space_refined.pdb - the resulting optimized model

## Build Up Complex / Run Complex on Phenix.real_space_refine

We have now quickly optimized a single subunit. Its stereochemistry and fit-to-density should have improved. Ideally, we want to take this one step further and optimized the interfaces. To do this we will build up the complex with the optimized subunit in Chimera.

1. Build up the complex in Chimera using **copy/combine…** in the model panel.

   a. Make three duplicates of the real space refined model.
   b. Fit each subunit into the 3 un-modeled subunit densities.
   c. Select all 4 subunits in the model panel and again use **copy/combine…** to join all 4 subunits creating one combined PDB file 4 chains.
   d. Save the new combined PDB file as 1DP0_chainA_rsr_complex.pdb. Be sure to save only the last combined model and save it relative to the map.
2. Run phenix.real_space_refine on the complex. Back to the terminal and type in the following command:

   ```
   Optimize hryc$ phenix.real_space_refine
   1DP0_chainA_rsr_complex.pdb emd_5995.map resolution=3.2
   run=minimization_global+adp ncs_constraints=True
   ```

   Here we are specifying the operations that will be run in the refinement (global minimization and compute atomic displacement parameters (ADP or B-factors per-atom)
3. During this run the user should note a few things:
   a. Four chains are being optimized (A,B,C,D).
   b. If one looks at the Start info that is output, the cross correlation (CC) around atoms is very similar to that of the unit cell. This is because we are now modeling the complete complex. Ideally these values will improve throughout the refinement.
   c. We made sure NCS constraints (On as default) was on since we would like to treat all the subunits identically. This reduces time during the refinement. Typically, if the subunits are not identical we turn off NCS constraints (ncs_constraints=False) at this step.
4. The output will take slightly longer than that of the initial model.
5. Again, the output will include three files.
6. Typically another refinement is run with ramachandran_restraints=False. Following this COOT is used to adjust regions with poor dihedral angles, improving the Ramachandran plot. This step is then iterated with clashscore, rotamers, and fit-to-density.

## Run Phenix.real_space_refine on **your** Reconstruction (if we have time)

The most difficult part of optimizing the model will be preparing the data.

1. In Chimera, open your reconstruction (or my reconstruction for ease of use - EMAN2_threed4.map) and the emd_5995.map.
2. If you open your own reconstruction, make sure the maps are aligned.
3. Resample your map onto the emd_5995.map grid. To do this open the Chimera command line and type "**vop resample #0 ongrid #1**", where #0 is your map and #1 is the emd_5995.map.

4. Open Volume Viewer (**Tools>Volume Data>Volume Viewer**) and click (**File>Save Map As…**) in the volume viewer.
5. Save the map as EMAN2_threedX.mrc where X is a resolution estimate so you don't forget when running Phenix.real_space_refine.
6. Close Chimera.
7. Open the terminal and navigate to the directory with your map and the real space refined complex. **Change the name of your complex from .mrc to .map** (Phenix likes .map files).
8. Run real space refinement on your reconstructed data and the modeled complex:

   ```
   Optimize hryc$ phenix.real_space_refine
   1DP0_chainA_rsr_complex_real_space_refined.pdb
   EMAN2_threed4.map resolution=4 run=minimization_global+adp
   ```

   Where **`EMAN2_threed4.map`** is my map reconstructed to 4Å and thus my stated resolution is 4Å in the command.
9. You have now optimized a model for your reconstruction!
10. Time to validate…

---

**Helpful References:**

https://www.phenix-online.org/documentation/reference/real_space_refine.html

Bartesaghi, Alberto, Doreen Matthies, Soojay Banerjee, Alan Merk, and Sriram Subramaniam. "Structure of B-galactosidase at 3.2-Å Resolution Obtained by Cryo-electron Microscopy." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 32 (2014): doi:10.1073/pnas.1402809111.

Wang, Zhao, Corey F Hryc, Benjamin Bammes, Pavel V Afonine, Joanita Jakana, Dong-Hua Chen, Xiangan Liu, *and others.* "An Atomic Model of Brome Mosaic Virus Using Direct Electron Detection and Real-space Optimization." *Nature communications* 5 (2014): doi:10.1038/ncomms5808.